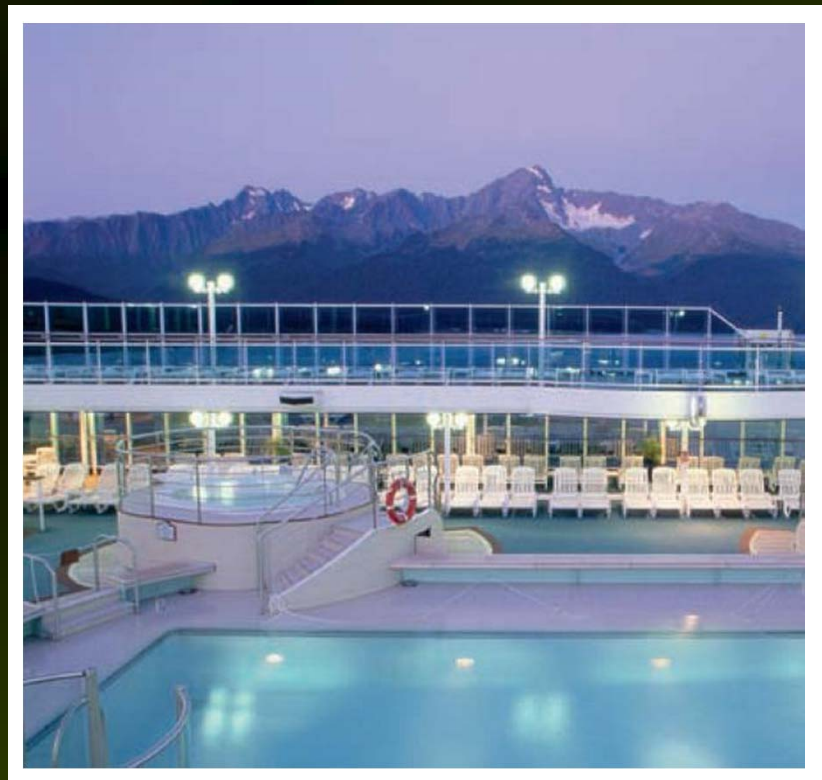
8

CALCULUS OF SEVERAL VARIABLES



8.4 The Method of Least Squares

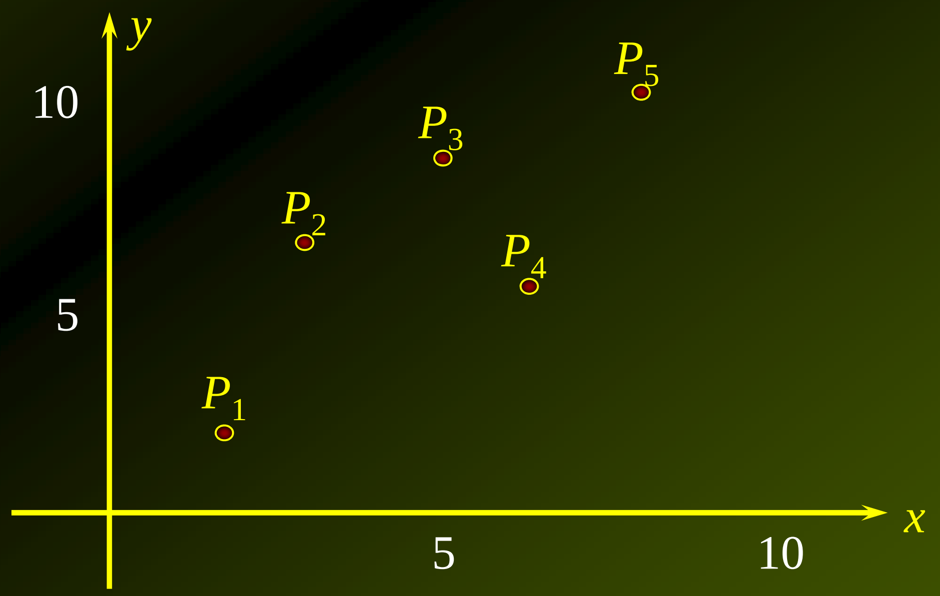
The Method of Least Squares

Suppose we are given the data points

$P_1(x_1, y_1)$, $P_2(x_2, y_2)$, $P_3(x_3, y_3)$, $P_4(x_4, y_4)$, and $P_5(x_5, y_5)$

that describe the relationship between two variables x and y.

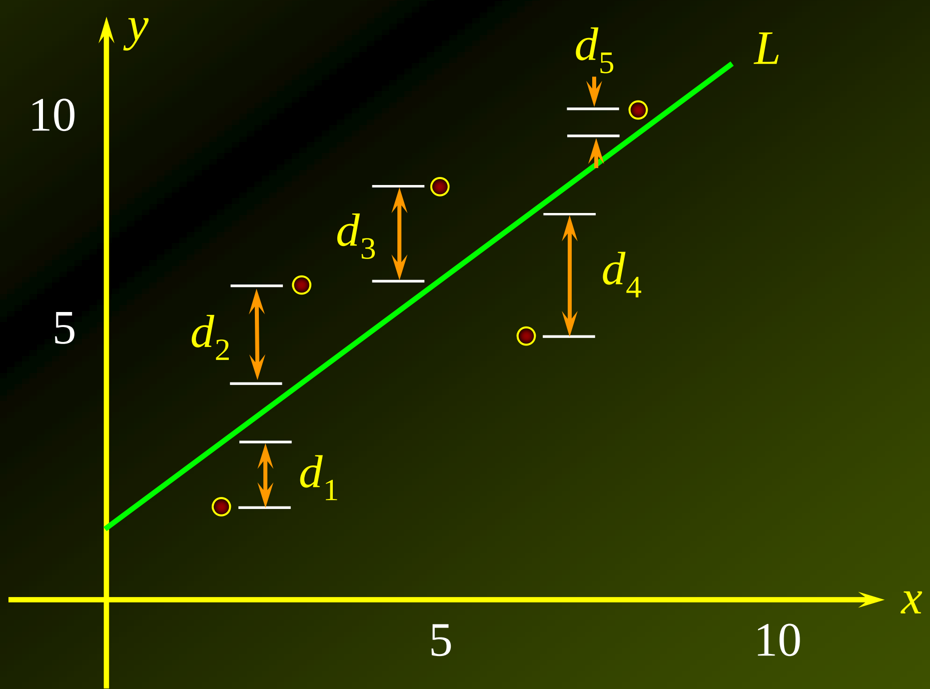
By plotting these data points, we obtain a scatter diagram:



The Method of Least Squares

Suppose we try to fit a straight line L to the data points P_1, P_2, P_3, P_4, and P_5.

The line will miss these points by the amounts d_1, d_2, d_3, d_4, and d_5 respectively.

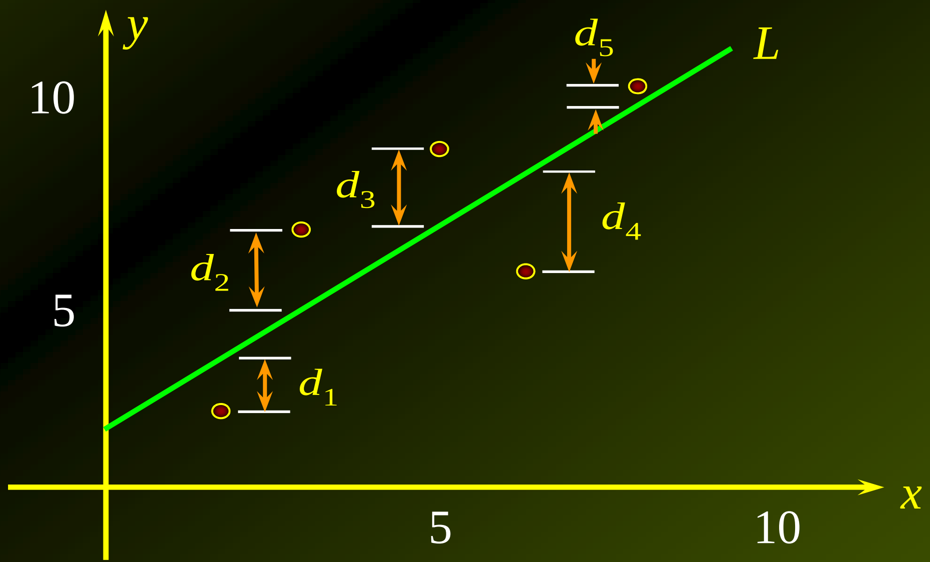


The Method of Least Squares

The principle of least squares states that the straight line L that fits the data points best is the one chosen by requiring that the sum of the squares of d_1, d_2, d_3, d_4, and d_5, that is

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

be made as small as possible.

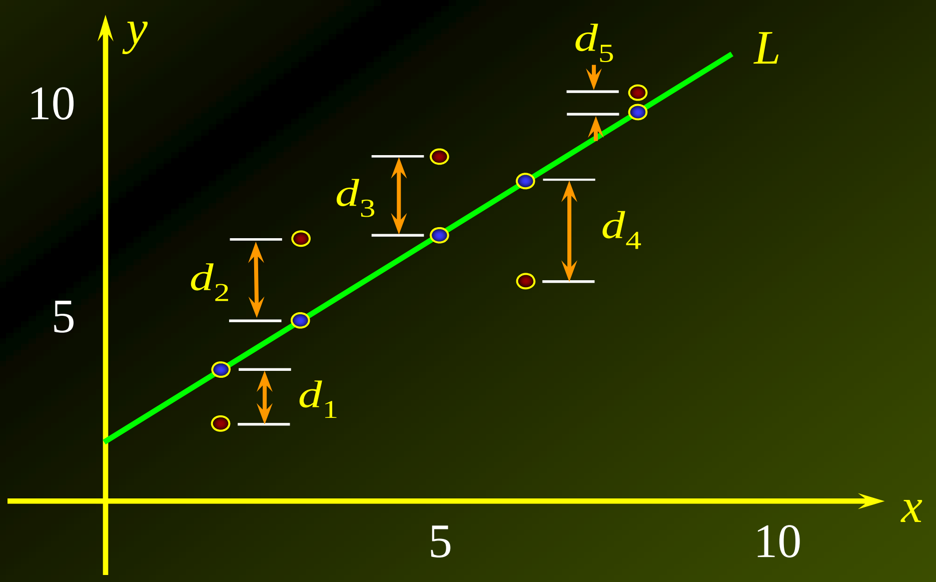


The Method of Least Squares

Suppose the regression line L is $y = f(x) = mx + b$, where m and b are to be determined.

The distances d_1, d_2, d_3, d_4, and d_5, represent the errors the line L is making in estimating these points, so that

$d_1 = f(x_1) - y_1,\ d_2 = f(x_2) - y_2,\ d_3 = f(x_3) - y_3,$ and so on.



The Method of Least Squares

Observe that

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

$$= [f(x_1) - y_1]^2 + [f(x_2) - y_2]^2 + [f(x_3) - y_3]^2 + [f(x_4) - y_4]^2 + [f(x_5) - y_5]^2$$

$$= [mx_1 + b - y_1]^2 + [mx_2 + b - y_2]^2 + [mx_3 + b - y_3]^2 + [mx_4 + b - y_4]^2 + [mx_5 + b - y_5]^2$$

This may be viewed as a function of two variables *m* and *b*.

The Method of Least Squares

Thus, the least-squares criterion is equivalent to minimizing the function

$$f(m,b) = (mx_1 + b - y_1)^2 + (mx_2 + b - y_2)^2 + (mx_3 + b - y_3)^2 + (mx_4 + b - y_4)^2 + (mx_5 + b - y_5)^2$$

We want to minimize

$$f(m,b) = (mx_1 + b - y_1)^2 + (mx_2 + b - y_2)^2 + (mx_3 + b - y_3)^2 + (mx_4 + b - y_4)^2 + (mx_5 + b - y_5)^2$$

We first find the partial derivative with respect to *m*:

$$\frac{\partial f}{\partial m} = 2(mx_1 + b - y_1)x_1 + 2(mx_2 + b - y_2)x_2 + 2(mx_3 + b - y_3)x_3 + 2(mx_4 + b - y_4)x_4 + 2(mx_5 + b - y_5)x_5$$

The Method of Least Squares

$$= 2[mx_1^2 + bx_1 - y_1x_1 + mx_2^2 + bx_2 - y_2x_2 + mx_3^2 + bx_3 - y_3x_3$$
$$+ mx_4^2 + bx_4 - y_4x_4 + mx_5^2 + bx_5 - y_5x_5]$$

$$= 2[(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2)m + (x_1 + x_2 + x_3 + x_4 + x_5)b$$
$$- (y_1x_1 + y_2x_2 + y_3x_3 + y_4x_4 + y_5x_5)]$$

We want to minimize

$$f(m,b) = (mx_1 + b - y_1)^2 + (mx_2 + b - y_2)^2 + (mx_3 + b - y_3)^2$$
$$+ (mx_4 + b - y_4)^2 + (mx_5 + b - y_5)^2$$

The Method of Least Squares

We now find the partial derivative with respect to *b*:

$$\frac{\partial f}{\partial b} = 2(mx_1 + b - y_1) + 2(mx_2 + b - y_2) + 2(mx_3 + b - y_3)$$

$$+ 2(mx_4 + b - y_4) + 2(mx_5 + b - y_5)$$

$$= 2[(x_1 + x_2 + x_3 + x_4 + x_5)m + 5b - (y_1 + y_2 + y_3 + y_4 + y_5)]$$

The Method of Least Squares

Setting

$$\frac{\partial f}{\partial m} = 0 \quad \text{and} \quad \frac{\partial f}{\partial b} = 0$$

gives

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2)m + (x_1 + x_2 + x_3 + x_4 + x_5)b$$
$$= y_1x_1 + y_2x_2 + y_3x_3 + y_4x_4 + y_5x_5$$

and

$$(x_1 + x_2 + x_3 + x_4 + x_5)m + 5b = y_1 + y_2 + y_3 + y_4 + y_5$$

Solving the two simultaneous equations for *m* and *b* then leads to an equation *y* = *mx* + *b*.

This equation will be the 'best fit' line, or regression line for the given data points.

The Method of Least Squares

Suppose we are given n data points:

$$P_1(x_1, y_1), P_2(x_2, y_2), P_3(x_3, y_3), \dots , P_n(x_n, y_n)$$

Then, the least-squares (regression) line for the data is given by the linear equation

$$y = f(x) = mx + b$$

where the constants m and b satisfy the equations

$$(x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2)m + (x_1 + x_2 + x_3 + \dots + x_n)b$$
$$= y_1x_1 + y_2x_2 + y_3x_3 + \dots + y_nx_n$$

and

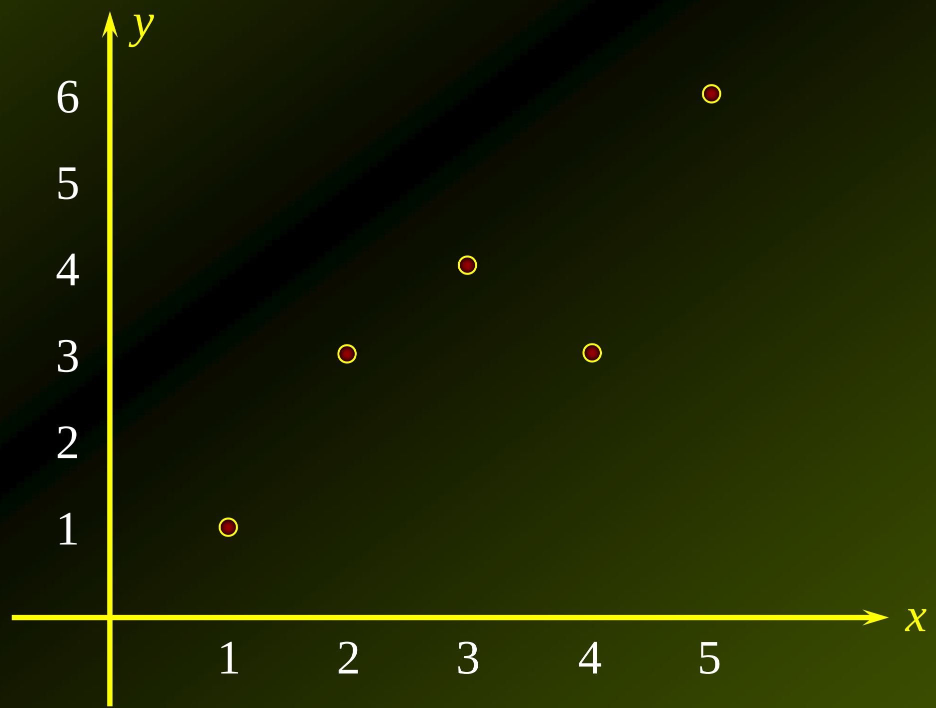
$$(x_1 + x_2 + x_3 + \dots + x_n)m + nb = y_1 + y_2 + y_3 + \dots + y_n$$

simultaneously.

These last two equations are called normal equations.

Example 1

Find the equation of the least-squares line for the data
$P_1(1, 1)$, $P_2(2, 3)$, $P_3(3, 4)$, $P_4(4, 3)$, and $P_5(5, 6)$



Example 1 – *Solution*

Here, we have $n = 5$ and

$$x_1 = 1 \quad x_2 = 2 \quad x_3 = 3 \quad x_4 = 4 \quad x_5 = 5$$
$$y_1 = 1 \quad y_2 = 3 \quad y_3 = 4 \quad y_4 = 3 \quad y_5 = 6$$

Substituting in the first equation we get

$$(x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2)m + (x_1 + x_2 + x_3 + \ldots + x_n)b$$
$$= y_1x_1 + y_2x_2 + y_3x_3 + \ldots + y_nx_n$$

$$(1^2 + 2^2 + 3^2 + 4^2 + 5^2)m + (1 + 2 + 3 + 4 + 5)b$$
$$= (1)(1) + (3)(2) + (4)(3) + (3)(4) + (6)(5)$$

$$55m + 15b = 61$$

Example 1 – *Solution*

cont'd

Substituting in the second equation we get

$$(x_1 + x_2 + x_3 + \ldots + x_n)m + 5b = y_1 + y_2 + y_3 + \ldots + y_n$$

$$(1 + 2 + 3 + 4 + 5)m + 5b = 1 + 3 + 4 + 3 + 6$$

$$15m + 5b = 17$$

Solving the simultaneous equations

$$55m + 15b = 61 \qquad 15m + 5b = 17$$

gives $m = 1$ and $b = 0.4$.

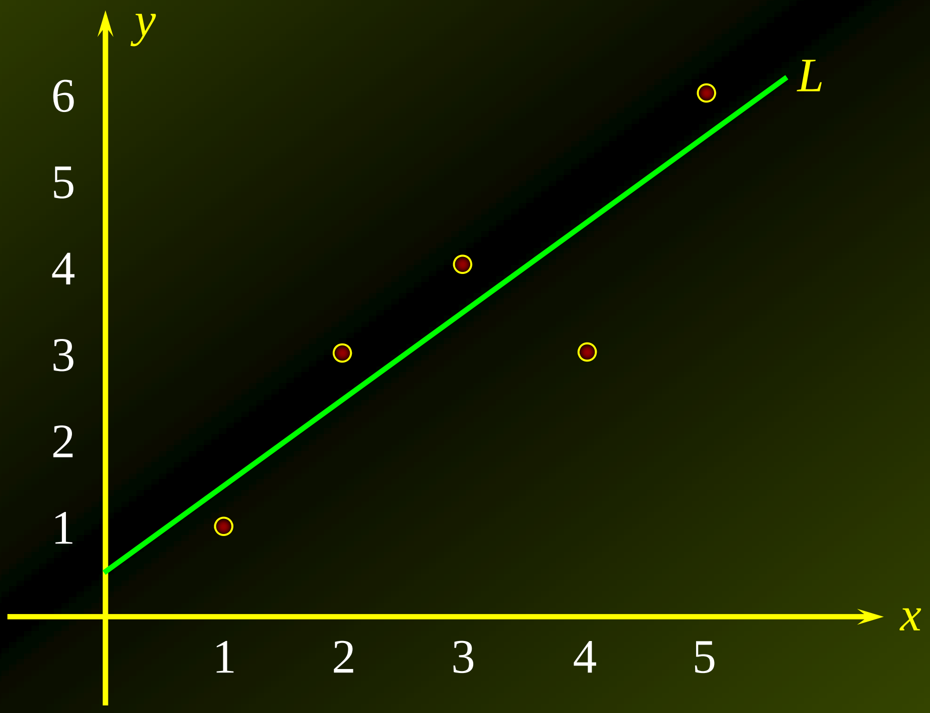
Therefore, the required least-squares line is

$$y = x + 0.4$$

Example 1 – *Solution*

cont'd

Below is the graph of the required least-squares line $y = x + 0.4$



Applied Example 3 – *Maximizing Profit*

A market research study provided the following data based on the projected monthly sales x (in thousands) of an adventure movie DVD.

p	38	36	34.5	30	28.5
x	2.2	5.4	7.0	11.5	14.6

a. Find the demand equation if the demand curve is the least-squares line for these data.

Applied Example 3 – *Maximizing Profit* cont'd

b. The total monthly cost function associated with producing and distributing the DVD is given by

$$C(x) = 4x + 25$$

where x denotes the number of discs (in thousands) produced and sold, and $C(x)$ is in thousands of dollars. Determine the unit wholesale price that will maximize monthly profits.

Applied Example 3(a) – *Solution*

The calculations required for obtaining the normal equations may be summarized as follows:

x	p	x^2	xp
2.2	38.0	4.84	83.6
5.4	36.0	29.16	194.4
7.0	34.5	49.00	241.5
11.5	30.0	132.25	345.0
14.6	28.5	213.16	416.1
40.7	167.0	428.41	1280.6

Applied Example 3(a) – *Solution*

cont'd

Thus, the nominal equations are

$$5b + 40.7m = 167 \quad \text{and} \quad 40.7b + 428.41m = 1280.6$$

Solving the system of linear equations simultaneously, we find that $m \approx -0.81$ and $b \approx 39.99$

Therefore, the required demand equation is given by

$$p = f(x) = -0.81x + 39.99 \qquad (0 \le x \le 49.37)$$

Applied Example 3(b) – *Solution*

cont'd

The total revenue function in this case is given by

$$R(x) = xp = x(-0.81x + 39.99)$$
$$= -0.81x^2 + 39.99x$$

Since the total cost function is

$C(x) = 4x + 25$

we see that the profit function is

$$P(x) = R - C$$
$$= -0.81x^2 + 39.99x - (4x + 25)$$
$$= -0.81x^2 + 35.99x - 25$$

Applied Example 3(b) – *Solution*

cont'd

To find the absolute maximum of $P(x)$ over the closed interval [0, 49.37], we compute

$$P'(x) = -1.62x + 35.99$$

Since $P(x) = 0$, we find that $x \approx 22.22$ as the only critical point of P.

Finally, from the table

x	0	22.22	49.37
$P(x)$	–25	374.78	– 222.47

we see that the optimal wholesale price is

$$p = -0.81(22.22) + 39.99 = 21.99$$

or $21.99 per disc.